EMBO reports

scientific report

Convergent evolution of gene networks by single-gene duplications in higher eukaryotes

Gregory D. Amoutzias, David L. Robertson, Stephen G. Oliver & Erich Bornberg-Bauer[+†]
School of Biological Sciences, The University of Manchester, Manchester, UK

By combining phylogenetic, proteomic and structural information, we have elucidated the evolutionary driving forces for the gene-regulatory interaction networks of basic helix–loop–helix transcription factors. We infer that recurrent events of single-gene duplication and domain rearrangement repeatedly gave rise to distinct networks with almost identical hub-based topologies, and multiple activators and repressors. We thus provide the first empirical evidence for scale-free protein networks emerging through single-gene duplications, the dominant importance of molecular modularity in the bottom-up construction of complex biological entities, and the convergent evolution of networks.

Keywords: convergent evolution; bHLH protein; protein interaction networks; gene duplication

EMBO reports advance online publication 13 February 2004;
doi:10.1038/sj.embor.7400096

INTRODUCTION

Explaining the evolution of complexity has been a challenge to darwinian theory since its conception. At the molecular level, biological complexity involves networks of ligand–protein, protein–protein and protein–nucleic acid interactions in metabolism, signal transduction, gene regulation, protein synthesis and so on. As organismal complexity increases, it has been observed that more control is required for the positive and negative regulation of genes such that complexity correlates with an increase in both the ratio and absolute number of transcription factors (Levine & Tjian, 2003).

Both theoretical studies and genome analyses (Wagner, 1994, 2003; Mendoza & Alvarez-Buylla, 1998) have been used to examine the evolution of complex genetic networks. The duplication of genes is the predominant mechanism for the generation of new members of a protein family and so is central to the evolution of complexity. A duplicated gene can result in redundancy if multiple proteins have the same or overlapping function. Alternatively, due to the reduced selective constraint on protein evolution, one of the copies of the duplicated gene can become nonfunctional or, more significantly, can acquire a new function (Ohno, 1970; Wagner, 2001). The duplication that increases the size of a network may occur either via single-gene duplication events or by duplication of genes on a large scale, including the entire genome (Wagner, 1994). The need for networks to remain stable and functional in the cellular environment after the duplication event(s) is thought to favour whole-genome duplication (Papp *et al*, 2003). In addition, recent investigations on domain combinations have suggested that domain rearrangements (intrusion, loss and differential spacing between domains) are quite frequent and more important than previously assumed (Apic *et al*, 2001), especially for regulatory proteins. This means that pathways and networks not only evolve by the basic principles of gene evolution (gene duplication/loss and point mutation) but also they adapt by rearrangement of selectively advantageous 'building blocks'.

The basic helix–loop–helix (bHLH) protein family comprises an ancient class of eukaryotic transcription factors that are found in fungi, plants and animals (Moore *et al*, 2000). Due to their homo- and heterodimerization abilities, they form a complex protein–protein interaction network. The bHLH family is believed to have expanded together with the appearance of multicellularity (Ledent *et al*, 2002). In unicellular eukaryotes, such as yeast, bHLH proteins are involved in the regulation of several metabolic pathways (Massari & Murre, 2000). In contrast, the bHLH proteins of metazoa are involved in regulating the cell cycle (Luscher, 2001), sensing environmental signals (Gu *et al*, 2000), and also in developmental processes (Massari & Murre, 2000).

The bHLH transcription factors are named after their highly conserved ~60-amino-acid-long domain that consists of a basic region followed by the helix–loop–helix motif, which comprises two amphipathic α-helices separated by a variable-length loop (Littlewood & Evan, 1995). Two bHLH proteins form a functional homo- or heterodimer (that is, a four-helix bundle) through their HLH domains. Additionally, the two basic regions are responsible for recognizing and binding a core hexanucleotide DNA sequence, such as the E-box (Brownlie *et al*, 1997). It is usual for bHLH proteins to include additional domains that are responsible for the activation or repression of target gene activity. In some phylogenetic groups, additional dimerization domains (for example, leucine zipper (LZ), PAS or 'orange' domains) are

2.205 Stopford Building, Oxford Road, School of Biological Sciences, The University of Manchester, Manchester M13 9PT, UK
[†]Bioinformatics Division, Institute of Botany, School of Biological Sciences, University of Münster, Schlossplatz 4, D4814P, Germany
[+]Corresponding author. Tel: +49 0251 83 21630; Fax: +49 0251 83 21631;
E-mail: ebb@uni-muenster.de

Received 27 August 2003; revised 8 December 2003; accepted 13 January 2004; published online 13 February 2004

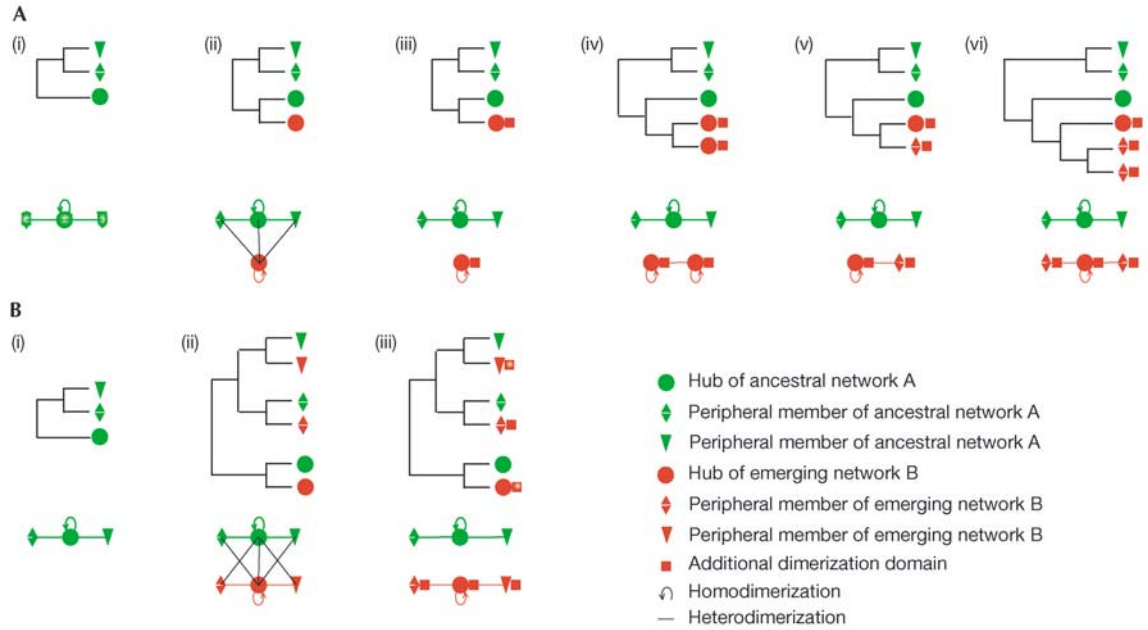
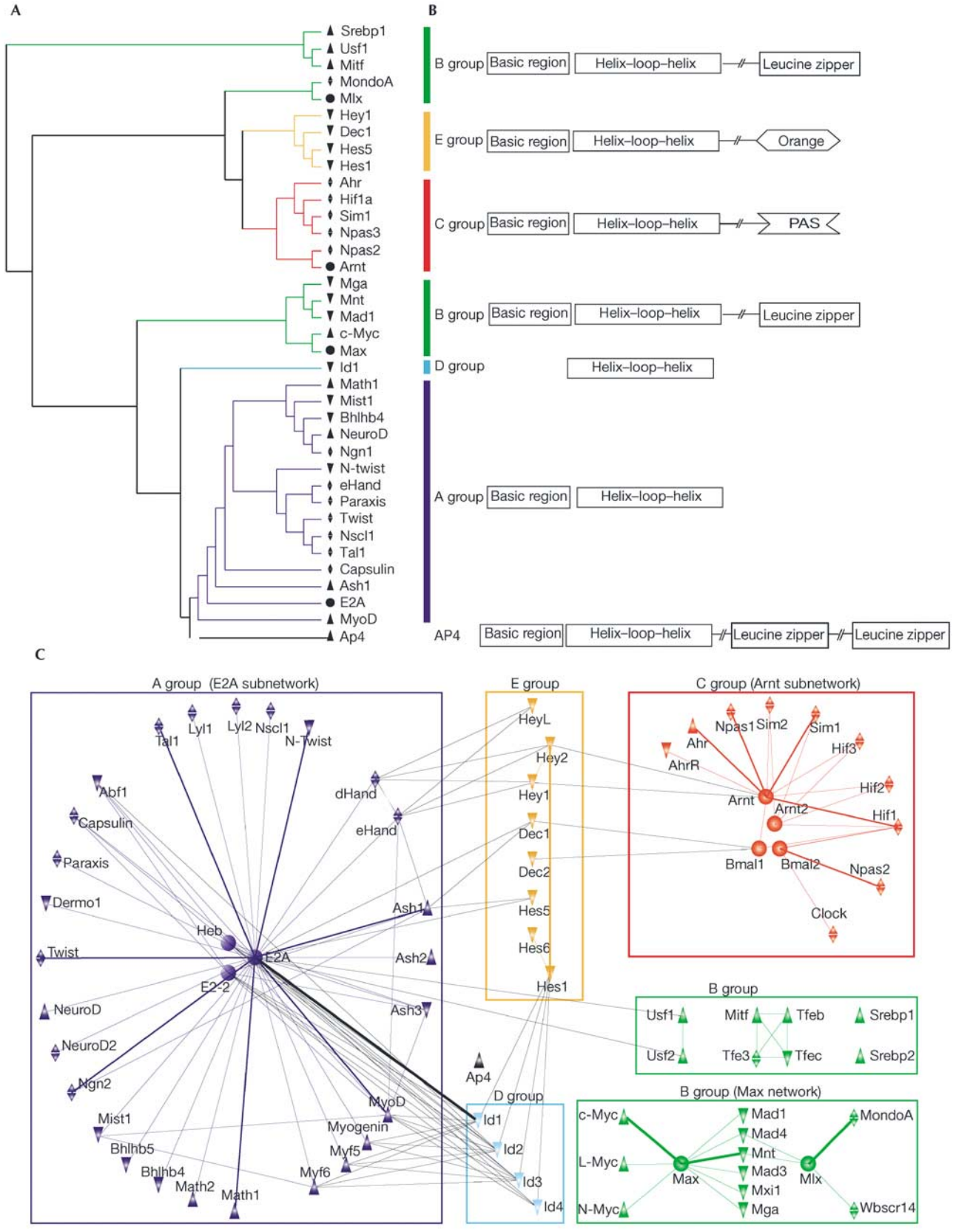


Fig 1 | Two possible patterns of network evolution. We denote hubs as circles, repressors as inverted triangles, factors that have ambivalent or ambiguous function as diamonds and additional dimerization domains as squares. (**A**) Evolution of a heterodimerization network by single-gene duplication. (i) Initial state of the ancestral network A. (ii) Single-gene duplication of the hub. The duplicated protein has the same dimerization properties as the ancestral hub. (iii) Intrusion of an additional dimerization domain in the emergent hub. Due to this intrusion, the emergent hub has higher affinity for itself (homodimer; indicated by an arrow) than the other members and, thus, is isolated from the ancestral network. (iv) Single-gene duplication of the emergent hub. The two new members can homodimerize as well as heterodimerize with each other. They have higher affinity for each other than members of the ancestral network. (v) Point mutations change the specificity of the newest member such that it can only heterodimerize with the emergent hub; therefore, it behaves as a peripheral member of the emergent network B. (vi) Single-gene duplication of the emergent peripheral member. The new protein has the same dimerization specificities as its parental protein. Both of them heterodimerize with the emergent hub. Members of the emergent network B have a monophyletic origin, and the additional dimerization domains have coevolved with the main dimerization domain. (**B**) Evolution of a heterodimerization network by large-scale gene duplication. (i) Initial state of the network. (ii) Large-scale gene duplication. Every duplicated gene will have the same dimerization specificities as its ancestral gene, thus forming a complex network. (iii) The duplicated members isolate from the ancestral network. The members of the emergent network are not monophyletic.

found at the C-terminal side of the HLH domain (Ledent & Vervoort, 2001).

In this paper, we investigate the evolution of a network of bHLH transcription factors and, while recent studies (Atchley & Fitch, 1997; Ledent & Vervoort, 2001; Ledent *et al*, 2002) have concentrated on sequence analysis, we have combined genomic data, domain architecture and protein–protein interactions. We have used the bHLH system to examine the question of whether single-gene or large-scale duplications had an important role in the evolution of this network. The conceptual differences and implications for phylogenetic analysis and the development of protein–protein interactions are illustrated in Fig 1. Recent work on genetic networks (Conant & Wagner, 2003), protein–protein interactions and protein complexes (Papp *et al*, 2003) has suggested that large-scale duplications are the major driving force for network evolution. The results in this report do not exclude a role for large-scale duplications. However, the main point that emerges from our study is the demonstration that recurrence of single-gene duplications and domain rearrangements repeatedly give rise to complex networks.

Fig 2 | bHLH heterodimerization network. (**A**) Cladogram of the human bHLH domains depicting a summary of the full neighbour joining tree (see supplementary information online), which is in accordance with previous publications (Atchley & Fitch, 1997; Ledent & Vervoort, 2001). (**B**) Domain architecture of the bHLH class of transcription factors, including the DNA-binding basic region, the HLH dimerization domain and other additional dimerization domains that lie C-terminal to the HLH. These additional dimerization domains are believed to confer dimerization specificity. (**C**) Topology of the bHLH protein network, based on the protein–protein interactions among members of the bHLH class. The network is compartmentalized according to five phylogenetic groups (A–E), based on both our own analyses and those of others (Atchley & Fitch, 1997; Ledent & Vervoort, 2001). We denote hubs as circles, activators as triangles, repressors as inverted triangles, and factors that have ambivalent or ambiguous function as diamonds. Interactions observed between *Drosophila* bHLH proteins have also been confirmed for mammalian proteins of the same subfamily. They are assumed to be ancestral and highly conserved, and therefore are denoted with thicker lines. ▶


A
Srebp1
Usf1
Mitf
MondoA
Mlx
Hey1
Dec1
Hes5
Hes1
Ahr
Hif1a
Sim1
Npas3
Npas2
Arnt
Mga
Mnt
Mad1
c-Myc
Max
Id1
Math1
Mist1
Bhlhb4
NeuroD
Ngn1
N-twist
eHand
Paraxis
Twist
Nscl1
Tal1
Capsulin
Ash1
E2A
MyoD
Ap4
B
B group Basic region Helix–loop–helix // Leucine zipper
E group Basic region Helix–loop–helix // Orange
C group Basic region Helix–loop–helix // PAS
B group Basic region Helix–loop–helix // Leucine zipper
D group Helix–loop–helix
A group Basic region Helix–loop–helix
AP4 Basic region Helix–loop–helix // Leucine zipper // Leucine zipper
C
A group (E2A subnetwork)
Tal1
Lyl1
Lyl2
Nscl1
N-Twist
Abf1
Capsulin
Paraxis
Dermo1
Twist
NeuroD
NeuroD2
Ngn2
Mist1
Bhlhb5
Bhlhb4
Math2
Math1
dHand
eHand
Ash1
Heb
E2A
E2-2
Ash2
Ash3
MyoD
Myogenin
Myf5
Myf6
E group
HeyL
Hey2
Hey1
Dec1
Dec2
Hes5
Hes6
Hes1
Ap4
D group
Id1
Id2
Id3
Id4
C group (Arnt subnetwork)
Npas1
Sim2
Sim1
Ahr
AhrR
Hif3
Hif2
Arnt
Arnt2
Hif1
Bmal1
Bmal2
Npas2
Clock
B group
Usf1
Mitf
Tfeb
Srebp1
Usf2
Tfe3
Tfec
Srebp2
B group (Max network)
c-Myc
L-Myc
N-Myc
Max
Mad1
Mad4
Mnt
Mad3
Mxi1
Mga
Mlx
MondoA
Wbscr14

RESULTS

Phylogenetic analyses based on the bHLH domain of the fungal and metazoan members of the bHLH protein family have classified them into four (A–D; Atchley & Fitch, 1997) or six (A–F; Ledent & Vervoort, 2001) major groups. The latter study split the B group of Atchley & Fitch (1997) into groups B and E, and also introduced the F group. The structural irregularities in the members of the F group raise considerable doubts about their classification as bHLH transcription factors, and so we excluded this group from our analysis. The short length of the bHLH domain, which consists of only ~60 amino acids, compromises the reliability of phylogenetic analysis. However, the overall tree topology is congruent with previous studies (Atchley & Fitch, 1997; Ledent & Vervoort, 2001) and strongly supported by domain architecture as each group has a distinct domain arrangement (Fig 2A,B), apart from group B, which is paraphyletic (Fig 2A; Ledent & Vervoort, 2001). As yeast and plant sequences are only found in the B group, the paraphyly is consistent with group B being ancestral to the other four groups.

When we combined the phylogenetic analysis with heterodimerization/protein–protein interaction data, two distinct hub-based networks became apparent (Fig 2C). Moreover, distinct subnetworks can be identified. Each of these subnetworks contains at least one hub, a bHLH protein that interacts with a large number of other poorly connected 'peripheral' proteins. Although all hubs also homodimerize, only a few of the peripheral members do. The criterion for splitting a network into smaller subnetworks is the presence of hubs and peripheral members of the same phylogenetic group and with similar domain architecture (Fig 2C). Here, we designate each of the hub-based networks and subnetworks by using the name of the protein or the family that acts as the hub: (1) the '*Max*' network in the B group, (2) the '*Arnt*' subnetwork in the C group and (3) the '*E2A*' subnetwork in the A group. Each hub-based network or subnetwork also has a distinct domain architecture (Fig 2B). The '*Arnt*' and '*E2A*' subnetworks are connected by the E group (HES family), together forming the '*E2A–Arnt*' network. Note that the four members of the Mitf family in the B group are all connected and form an independent network, but as it is not hub based, they are not considered further.

Mathematical analysis confirms that the structure of these networks is nonrandom and hub based following the scale-free principle (see supplementary information online). This means that the distribution of connectivity decays as a power law $P(k) \sim k^{-\gamma}$, where k is the number of connections of a node and $P(k)$ is the frequency of nodes with k interactions (Barabasi, 2002). Interestingly, the relative connectivity of the hubs in bHLH networks is higher than that found in the other biological networks that have been analysed. In terms of network theory, this is expressed as a lower γ value in the power-law equation for the bHLH network (where $\gamma \sim 1$) than in most other networks (where γ ranges from 2 to 3; Goh *et al*, 2002). This means that there are fewer connections between the peripheral members of the bHLH network than in other networks. As it is conceivable that the higher connectivity of the bHLH hubs results from a bias in the data, we examined the literature carefully to check this result. Although we found a significant number of reports in which peripheral members were tested explicitly, they either exhibited a very restricted range of interactions or no interactions at all with other peripheral members (Hogenesch *et al*, 1997; Firulli *et al*, 2000; Luscher, 2001). We also note that most bHLH interactions have been confirmed by more than one independent method (see supplementary information online). The high connectivity of hubs in the bHLH networks appears to be a direct consequence of the fact that gene duplication events (single or large scale) have generated new peripheral proteins that then interact preferentially with the hub.

The hub proteins in the bHLH protein-interaction networks are usually widely expressed in different tissues and organs. They need to heterodimerize with peripheral members of the network, which have a more limited expression pattern, to exert their different effects. The peripheral members of the networks are either activators or inhibitors of transcription, and their formation of a heterodimer with the hub protein usually allows them to form a functional complex that binds specific promoter elements (for example, E-boxes). In several cases, dimerization may occur among peripheral members of the network alone, but these dimers are usually non-DNA-binding and therefore repress transcription.

The '*Max*' network is involved in cell cycle control and it has no known bHLH protein interaction with the '*E2A–Arnt*' network; this could be due to the nature of the additional dimerization domain (LZ) that seems to impose specificity on the range of bHLH interactions (Bornberg-Bauer *et al*, 1998). The members of the '*Arnt*' subnetwork function mainly as environmental sensors controlling molecular clocks, the hypoxic response and the metabolism of toxic substances such as arylhydrocarbons. Finally, the '*E2A*' subnetwork is involved in developmental processes.

Interestingly, the topology of the '*Max*' network parallels that of the '*E2A–Arnt*' network (see supplementary information online). In '*Max*', there are two hubs (Max and Mlx) with peripheral partners that are either activators or repressors. The two hubs are connected through the Mad family, which act as repressor proteins. In the '*E2A–Arnt*' network, there are also two hub families (Arnt and E2A families), and their peripheral partners are either activators or repressors. As with '*Max*', repressor proteins (this time, of the HES family) connect these two hub families within the '*E2A–Arnt*' network.

The most striking finding is that the parallelism between the '*Max*' and the '*E2A–Arnt*' networks extends to their phylogenetic relationships (Fig 2A). In the '*Max*' network, the two hubs are not clustered together phylogenetically, which argues against their evolution by genome duplication or other large-scale events. The 'bridge', that is, the molecules that link the two hubs, is a family of repressors that are phylogenetically quite distant from the hubs. The same pattern appears in the '*E2A–Arnt*' network. The two hub families (Arnt and E2A) are not clustered together in the sequence-based phylogeny, but (instead) cluster with their respective activators or repressors. Again, the 'bridge' between them is a family of repressors (the HES family) that are phylogenetically quite distant from the hubs. All of this suggests that there are similar restraints on the evolution of the different networks in the bHLH protein family. Newly emergent networks and subnetworks have followed this same evolutionary process at least twice. Thus, these bHLH networks show evolutionary convergence, giving rise to a symmetrical topology. The hub proteins have a key role in this convergence, as they act as central regulators that enable appropriate choices to be made between alternate patterns of gene expression.

Further evidence for a common mechanism of network evolution comes from another interesting parallelism between the '*Arnt*' and '*E2A*' subnetworks. In each, there exists a family of

proteins (Period, which is not included in our analysis, and Id, respectively) that can bind with the hub or some peripheral members. As Period and Id lack either the bHLH or the basic region domain, they form nonfunctional heterodimers that cannot bind E-boxes. Thus, in both the '*E2A*' and '*Arnt*' subnetworks, there exists a mechanism that functionally sequesters the hubs or the activators by forming nonfunctional heterodimers. In other words, the same inhibitory mechanism is used on both 'sides' of the network.

Many of the binding specificities and their evolution can be understood from a structural perspective. For example, the inclusion of the PAS domain in group C restricts the range of interactions in which the bHLH domain may participate (Pongratz *et al*, 1998). Moreover, in the case of the AP-4 protein, the presence of two LZ domains inhibits its dimerization with the E2A hub (Hu *et al*, 1990). Furthermore, the differential spacing of the LZ from the HLH domain in the B group has been regarded as a mechanism of restricting interactions in the TFE3 factor (Beckmann & Kadesch, 1991). It is also well known that a small number of point mutations may cause subtle structural alterations in the surface of dimerization domains that result in significant changes in dimerization specificity (LZs of the '*Max*' network) (Nair & Burley, 2003). These findings increase our confidence that many peripheral members do, in fact, have only a limited number of interactions and that this result is not due to experimental bias.

DISCUSSION

There are several conclusions from our analysis. First, our results suggest that, for the evolution of networks based on one kind of binding domain (such as the bHLH), a model of single-gene duplication followed by domain rearrangements, point mutations and ongoing gene duplication is sufficient to generate quite complex interaction patterns, which mediate activation and repression. This finding does not preclude a role for large-scale gene duplication. For example, the postulated two rounds of whole-genome duplication in early vertebrates (Ohno, 1970) increased the complexity of networks by increasing the number of paralogous genes in each family. Nevertheless, our results strongly suggest that it is via single-gene duplications and domain rearrangements that new networks first arose in early metazoan evolution, before the divergence of arthropods and chordates. The mechanism of frequent gene duplication and domain rearrangement has also been demonstrated for signal transduction proteins involved in metazoan development (Miyata & Suga, 2001). To our knowledge, the emergence of a hub-based network with scale-free properties has not been explicitly explained before using phylogenies based on real data. This mechanism of evolution could be extrapolated to other families of transcription factors or signal transduction proteins that form complex dimerization networks and expanded during early metazoan evolution.

Second, we find a compelling symmetry between the two networks ('*Max*' and '*E2A–Arnt*') as well as similar structures within them. The '*E2A–Arnt*' network resembles an expanded and duplicated version of the '*Max*' network. However, the phylogenetic analysis does not support large-scale gene or genome duplication as a mechanism for the generation of the two networks. Rather, our results indicate that each of the networks evolved independently by single-gene duplication events towards a similar topology (in terms of hubs, repressors and activators), thus providing a striking example of convergent evolution at the level of protein networks. The reason for this reliance on single-gene duplications is that, if a new network was generated by large-scale gene duplication, then all members of that network would have to be isolated from the 'parental' network. This is highly unlikely, given that we know that network specificity is defined by domain architecture (Hu *et al*, 1990; Bornberg-Bauer *et al*, 1998); thus, after entire network duplication, it would be necessary for the new architecture to have arisen simultaneously in all members of the new network (Fig 1). Accordingly, new domain architecture is only significant if it is initially duplicated in a singular fashion such that a new and isolated network forms.

In conclusion, the initiation of the evolution of new subnetworks by gene duplication in pre-existing networks, coupled with domain rearrangements and point mutations, provides the possibility for a multitude of new protein–protein interaction pairs to develop. We infer that, in the case of bHLH, early duplication events of genes ancestral to a particular network were followed by domain intrusions and losses. Although the picture might be different with other gene families, it appears that single-gene duplications cannot be dismissed as a feasible mechanism to generate complex networks, and that such duplications have facilitated the evolution of complexity.

MATERIAL AND METHODS

From 125 bHLH proteins (Ledent *et al*, 2002), 78 were initially chosen as they were well documented in the literature. Another 135 were identified using family-specific hidden Markov models. These were constructed from TRANSFAC (Matys *et al*, 2003) and used to scan the translated open reading frames of the selected organisms from NCBI (http://www.ncbi.nih.gov).

All protein interactions were confirmed by a comprehensive literature search (see supplementary information online). A list of 78 mammalian genes connected by 127 interactions was retrieved. Of these interactions, 17 have been verified for the same subfamilies among *Drosophila* proteins and are highlighted in Fig 2C with thicker lines.

Phylogenetic relationships were inferred by the neighbour joining method with the PHYLIP package (available at http://evolution.genetics.washington.edu/phylip.html).

The complete list of bHLH interactions, the full phylogenetic tree and the supplementary information can be viewed at http://www.bioinf.man.ac.uk/~amoutzias/bHLH-evolution.html.

Supplementary information is available at *EMBO reports* online (http://www.emboreports.org).

ACKNOWLEDGEMENTS

We thank Dr Cendrine Hudelot for helpful comments and discussion. G.D.A. also thanks Dimitris and Vasiliki Amoutzias for their support throughout this project. G.D.A. is the recipient of a CASE studentship from the EPSRC and AstraZeneca plc. E.B.-B. acknowledges support through an MRC international recruitment grant.

REFERENCES

Apic G, Gough J, Teichmann SA (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J Mol Biol* **310:** 311–325

Atchley WR, Fitch WM (1997) A natural classification of the basic helix–loop–helix class of transcription factors. *Proc Natl Acad Sci USA* **94:** 5172–5176

Barabasi A (2002) *Linked, the New Science of Networks*. Oxford: Perseus Press

Beckmann H, Kadesch T (1991) The leucine zipper of TFE3 dictates helix–loop–helix dimerization specificity. *Genes Dev* **5:** 1057–1066

Bornberg-Bauer E, Rivals E, Vingron M (1998) Computational approaches to identify leucine zippers. *Nucleic Acids Res* **26:** 2740–2746

Brownlie P, Ceska T, Lamers M, Romier C, Stier G, Teo H, Suck D (1997) The crystal structure of an intact human Max–DNA complex: new insights into mechanisms of transcriptional control. *Structure* **5:** 509–520

Conant C, Wagner A (2003) Convergent evolution of gene circuits. *Nat Genet* **34:** 264–266

Firulli BA, Hadzic DB, McDaid JR, Firulli AB (2000) The basic helix–loop–helix transcription factors dHAND and eHAND exhibit dimerization characteristics that suggest complex regulation of function. *J Biol Chem* **275:** 33567–33573

Goh KI, Oh E, Jeong H, Kahng B, Kim D (2002) Classification of scale-free networks. *Proc Natl Acad Sci USA* **99:** 12583–12588

Gu YZ, Hogenesch JB, Bradfield CA (2000) The PAS superfamily: sensors of environmental and developmental signals. *Annu Rev Pharmacol Toxicol* **40:** 519–561

Hogenesch JB, Chan WK, Jackiw VH, Brown RC, Gu YZ, Pray-Grant M, Perdew GH, Bradfield CA (1997) Characterization of a subset of the basic–helix–loop–helix–PAS superfamily that interacts with components of the dioxin signaling pathway. *J Biol Chem* **272:** 8581–8593

Hu YF, Luscher B, Admon A, Mermod N, Tjian R (1990) Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. *Genes Dev* **4:** 1741–1752

Ledent V, Vervoort M (2001) The basic helix–loop–helix protein family: comparative genomics and phylogenetic analysis. *Genome Res* **11:** 754–770

Ledent V, Paquet O, Vervoort M (2002) Phylogenetic analysis of the human basic helix–loop–helix proteins. *Genome Biol* **3:** RESEARCH0030

Levine M, Tjian R (2003) Transcription regulation and animal diversity. *Nature* **424:** 147–151

Littlewood TD, Evan GI (1995) Transcription factors 2: helix–loop–helix. *Protein Profile* **2:** 621–702

Luscher B (2001) Function and regulation of the transcription factors of the Myc/Max/Mad network. *Gene* **277:** 1–14

Massari ME, Murre C (2000) Helix–loop–helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol* **20:** 429–440

Matys V *et al* (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* **31:** 374–378

Mendoza L, Alvarez-Buylla ER (1998) Dynamics of the genetic regulatory network for *Arabidopsis thaliana* flower morphogenesis. *J Theor Biol* **193:** 307–319

Miyata T, Suga H (2001) Divergence pattern of animal gene families and relationship with the Cambrian explosion. *BioEssays* **23:** 1018–1027

Moore AW, Barbel S, Jan LY, Jan YN (2000) A genomewide survey of basic helix–loop–helix factors in *Drosophila*. *Proc Natl Acad Sci USA* **97:** 10436–10441

Nair SK, Burley SK (2003) X-ray structures of Myc–Max and Mad–Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* **112:** 193–205

Ohno S (1970) *Evolution by Gene Duplication*. New York: Springer

Papp B, Pal C, Hurst L (2003) Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424:** 194–197

Pongratz I, Antonsson C, Whitelaw ML, Poellinger L (1998) Role of the PAS domain in regulation of dimerization and DNA binding specificity of the dioxin receptor. *Mol Cell Biol* **18:** 4079–4088

Wagner A (1994) Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. *Proc Natl Acad Sci USA* **91:** 4387–4391

Wagner A (2001) Birth and death of duplicated genes in completely sequenced eukaryotes. *Trends Genet* **17:** 237–239

Wagner A (2003) How the global structure of protein interaction networks evolves. *Proc R Soc Lond B* **270:** 457–466